[lustre1.6]

Lustre 1.6 的安装配置手册

Make：王日友 (海参)
Time：2007-2-10 周 6
msn：wangriyou@hotmail.com

Part1 介绍
Part2 架构
Part3 安装
Part4 调试
Part5 总结

Part 1. 介绍

Lustre 介绍， 先略…

Part 2. 架构

Lustre 一个 Lustre 集群由下面主体三类系统组成 –

i. The Meta Data Server (MDS) ：元数据服务器
ii. The Object Storage Servers (OSS) ：对象存储服务器
iii. The Clients ：客户端

为了提供高可用(HA)，可以架设 MDS，OSS 位冗余系统。Failovre 使得容错转移，这样可以提高 Uptime 时间。现在的系统架构是为了实现 lustre 集群，搭建在虚拟机(vmware)，平台：RHEL-AS4-Up2(x86-32bit)，因为是测试系统操作系统选择完全安装。 实际需求软件： perl 支持，RHEL-AS4－Up2 的默认安装 perl 版本是 v5.8.5， 是可以满足需要的； 还需要 python（版本大于 2）在安装开发包的时候，是默认安装的，如果没有安装的话， 可以在 OS 安装 disk 上找到 rpm 包； PyXML 也是另外需要安装的支持包，系统盘上没有的， 需要自己在网络上找到 tar 包，编译好，安装即可，在这样的环境下不会出问题。 XML processor for python: requires PyXML 点击这里： http://sourceforge.net/project/showfiles.php?group_id=6473 ，

另外说一下，PyXML 的安装不像一般 tar 包的编译/安装（make && make install）。 不过可以看看 README，你能得到答案。

由于试验环境有限，整个架构节点少。所以架构简单；
系统环境：
3 节点架构， 一台客户端（CLS），一台元数据服务器（MDS），一台对象存储服务器（OSS）；
操作系统， 3 台机器操作系统统一（Red Hat Enterprise Linux AS4 UP2，2.6.9-22.EL），选择完全安装；
Luster 环境里面有客户端担当时间同步服务器， 其他机器作为客户端；
3 台服务器全是在 vmware 上的虚拟操作系统；
客户端一块虚拟硬盘（10GB）， 其他机器均有 2 块虚拟硬盘（分别 20GB）；
以上就是全部的测试环境， 下面是拓扑架构：

Luster 对系统的架构预计定义：

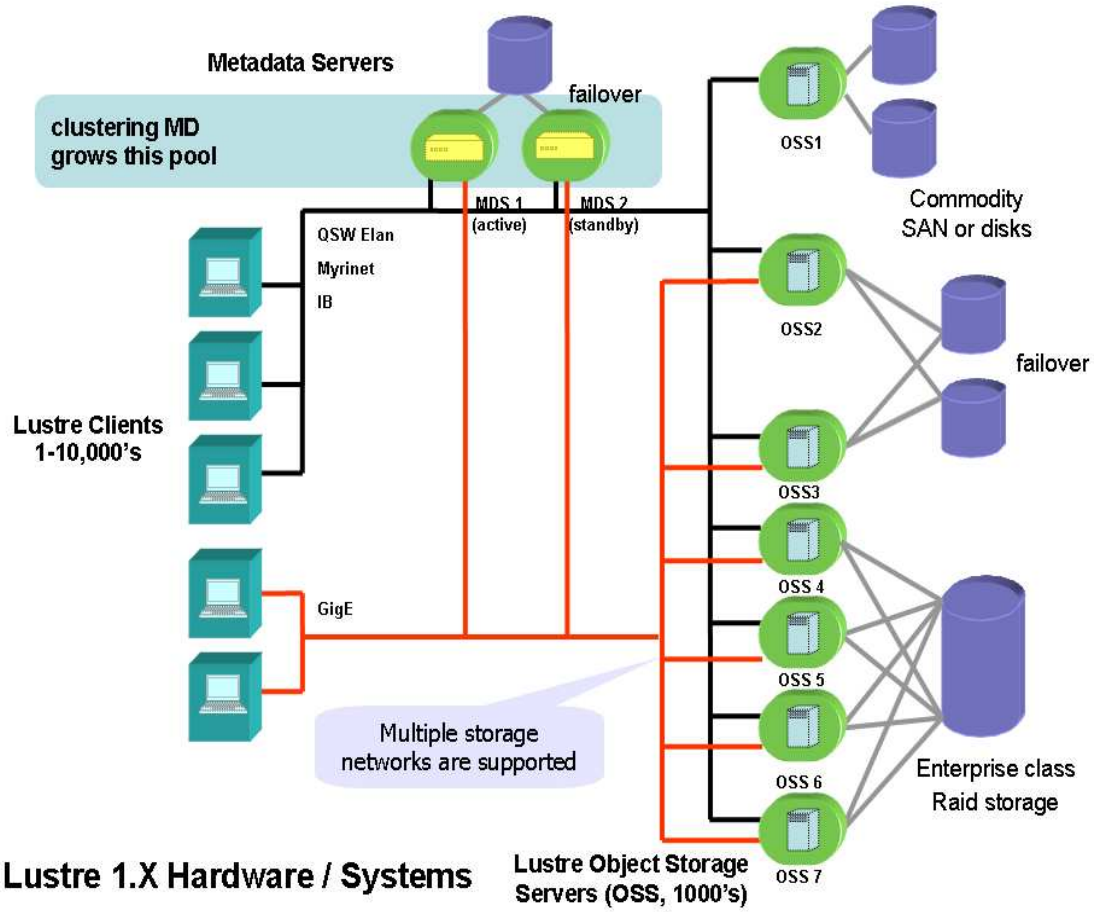

Figure 1.1 A Lustre cluster

实际试验系统架构：

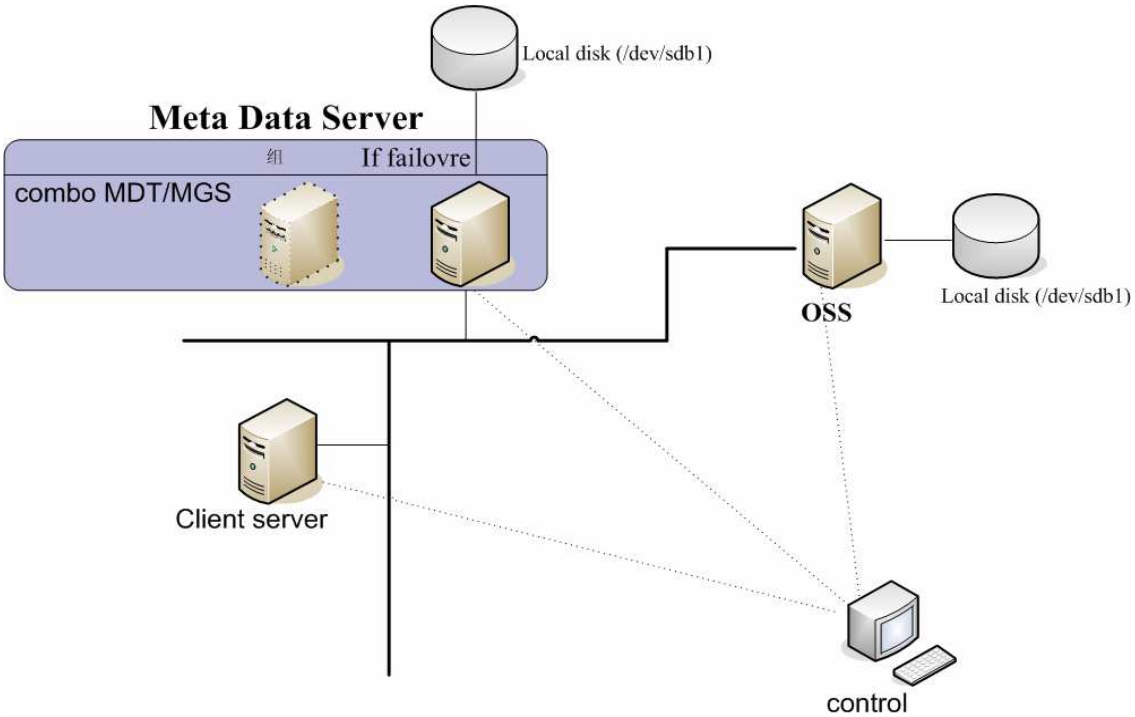

Figure 1.2 A Lustre cluster

Part 3. 安装

Lustre 安装过程， 这部分基本上是 linux 命令部分。 均是以 root 用户执行的所有操作。

- 软件安装

确认系统支持：

确认所有操作系统的版本，内核版本，在所有节点上执行下面命令

```
[root@CLS ~]# cat /etc/redhat-release

Red Hat Enterprise Linux AS release 4 (Nahant Update 2)

[root@CLS ~]# uname -a

Linux CLS 2.6.9-42.0.3.EL #1 SMP Thu Jan 11 17:55:05 MST 2007 i686 i686 i386 GNU/Linux
```

确认所有系统软件支持，软件版本， 在所有节点上执行下面命令

```
[root@CLS ~]# perl -v

This is perl, v5.8.5 built for i386-linux-thread-multi

[root@CLS ~]# python —V

Python 2.3.4 (#1, Feb  2 2005, 11:44:49)

[GCC 3.4.3 20041212 (Red Hat 3.4.3-9.EL4)] on linux2
```

安装必要的 lustre 包

下载 e2fsprogs 包，版本最好是大于 1.38。 下载地址：ftp://ftp.lustre.org/pub/lustre/other/e2fsprogs/

下载： e2fsprogs-devel-1.38.cfs1-0.i386.rpm， e2fsprogs-1.38.cfs1-0.i386.rpm

```
[root@CLS ~]# rpm -Uvh -nodeps e2fsprogs-devel-1.38.cfs1-0.i386.rpm

[root@CLS ~]# rpm -Uvh e2fsprogs-1.38.cfs1-0.i386.rpm
```

安装 PyXML

下载 PyXML, PyXML 点击这里： http://sourceforge.net/project/showfiles.php?group_id=6473

```
[root@CLS ~]# tar -zxvf PyXML-0.8.4.tar.gz

[root@CLS lustre]# cd PyXML-0.8.4

[root@CLS PyXML-0.8.4]# python setup.py build

[root@CLS PyXML-0.8.4]# python setup.py install
```

安装 lustre

安装 lustre 相关 rpm 包，本次测试安装 1.60 版本，因为有先后顺序。 所以需要安装下面顺序安装。可以在 lustre 网站下载相应版本， 不过现在 1.60 版是 beta 版本。

```
[root@CLS ~]# rpm -ivh  kernel-lustre-smp-2.6.9-42.0.3.EL_lustre.1.5.97.i686.rpm

[root@CLS ~]# rpm -ivh  kernel-lustre-source-2.6.9-42.0.3.EL_lustre.1.5.97.i686.rpm

[root@CLS ~]# rpm -ivh lustre-modules-1.5.97-2.6.9_42.0.3.EL_lustre.1.5.97smp.i686.rpm

[root@CLS ~]# rpm -ivh lustre-1.5.97-2.6.9_42.0.3.EL_lustre.1.5.97smp.i686.rpm

[root@CLS ~]# rpm -ivh lustre-source-1.5.97-2.6.9_42.0.3.EL_lustre.1.5.97smp.i686.rpm

[root@CLS ~]# rpm -ivh \
lustre-debuginfo-1.5.97-2.6.9_42.0.3.EL_lustre.1.5.97smp.i686.rpm
```

NODE: 以上操作在每一个节点都执行， 这样节点系统就全部安装了 luster 软件。

- 系统设置

模块配置

首要要设置 network 模块，在配置文件添加 lnet 的定义

```
[root@CLS ~]# vi /etc/modprobe.conf

alias eth0 pcnet32

alias eth1 pcnet32

alias scsi_hostadapter mptbase

alias scsi_hostadapter1 mptscsih

# add by wangriyou, Module options for networking

options lnet networks=tcp
```

域名问题

如果有域名对系统的路由将会有另外的对应设置， 目前本次测试没有。所以在每一个系统配置 hosts 文件。目前系统有 3 台服务器， 客户端 1 （CLS）， MD 服务器 1 （MDS） ， OS 服务器 1 (OBS)。

```
[root@CLS ~]# vi /etc/hosts

127.0.0.1            localhost.localdomain localhost

192.168.0.161  CLS
```

```
192.168.0.163  MDS

192.168.0.165  OST
```

在客户端服务器上执行

```
[root@CLS ~]# vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=CLS

[root@CLS ~]# hostname CLS
```

在 MD 服务器上执行

```
[root@CLS ~]# vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=MDS

[root@CLS ~]# hostname MDS
```

在 OS 服务器上执行

```
[root@CLS ~]# vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=OSS

[root@CLS ~]# hostname OSS
```

启动选项

编修 grub.conf 文件，使得启动 lustre 补丁内核

```
[root@CLS ~]# vi /boot/grub/grub.conf

# grub.conf generated by anaconda

# Note that you do not have to rerun grub after making changes to this file

# NOTICE:  You have a /boot partition.  This means that

#         all kernel and initrd paths are relative to /boot/, eg.

#         root (hd0,0)

#         kernel /vmlinuz-version ro root=/dev/sda3
```

```
#         initrd /initrd-version.img

#boot=/dev/sda

default=0

timeout=5

splashimage=(hd0,0)/grub/splash.xpm.gz

hiddenmenu

title Red Hat Enterprise Linux AS (2.6.9-42.0.3.EL_lustre.1.5.97smp)

        root (hd0,0)

        kernel /vmlinuz-2.6.9-42.0.3.EL_lustre.1.5.97smp ro root=LABEL=/1 rhgb quiet

        initrd /initrd-2.6.9-42.0.3.EL_lustre.1.5.97smp.img

title Red Hat Enterprise Linux AS (2.6.9-22.EL)

        root (hd0,0)

        kernel /vmlinuz-2.6.9-22.EL ro root=LABEL=/1 rhgb quiet

        initrd /initrd-2.6.9-22.EL.img
```

现在就可以重启动系统了。

时间同步

以 **CLS** 为时间同步服务器， 其他服务器为同步客户端。 也可以在本地设立时间同步服务器，这里就省略… …

Part 4. 调试

等全部 lustre 架构成员系统启动后，测试 hosts 文件的域名映射是否通信正常。 可以用 ping 工具来测试。如果不能做到互相通信，请确认问题。

创建并且启动文件系统

在 MD 服务器上执行以下操作

```
mkfs.lustre --fsname=testfs --mdt --mgs /dev/sdb1

mkdir -p /mnt/test/mdt
```

```
mount -t lustre /dev/sda1 /mnt/test/mdt

cat /proc/fs/lustre/devices

 0 UP mgs MGS MGS 5

 1 UP mgc MGC192.168.0.163@tcp e51f834d-c2e3-4c27-806b-cb66b7807420 5

 2 UP mdt MDS MDS_uuid 3

 3 UP lov testfs-mdtlov testfs-mdtlov_UUID 4

 4 UP mds testfs-MDT0000 testfs-MDT0000_UUID 3
```

在 OS 服务器上执行以下操作

```
mkfs.lustre --fsname=testfs --ost --mgsnode=MDS@tcp0 /dev/sdb2

mkdir -p /mnt/test/ost0

mount -t lustre /dev/sdb2 /mnt/test/ost0
```

在客户端服务器上执行一下操作

```
mkdir -p /mnt/testfs

mount -t lustre MDS@tcp0:/testfs /mnt/testfs

mount

192.168.0.163@tcp:/testfs on /mnt/testfs type lustre (rw)
```

可以看到客户端已经挂载了 lustre 文件系统， 也可以同样用 OS 服务器挂载。 在客户端上 testfs 文件系统上创建 test.txt 文件， 在 OS 服务器同时能看到 test.txt 文件的创建。

停止 lustre

```
[root@CLS ~]# umount  /mnt/testfs
```

Part 5. 总结

Luster1.6 还有很多 bug 与问题，但是相对以前的版本有很多性能和功能上的变革。 安装 1.6 前版本安装配置很复杂， 要用到旧有的 xml 的 schema。 在 1.6 配置相对简洁。 Good luck!